

Screening for inhibitors of main protease in SARS-CoV-2: *in silico* and *in vitro* approach avoiding secondary amides

Kazuki Z Yamamoto,[†] Nobuaki Yasuo,[‡] and Masakazu Sekijima^{*,†}

[†]*Department of Computer science, Tokyo Institute of Technology, Yokohama, 226-8501, Japan*

[‡]*Academy for Convergence of Materials and Informatics, Tokyo Institute of Technology, Tokyo, 152-8550, Japan*

E-mail: sekijima@c.titech.ac.jp

Phone: +123 (0)123 4445556. Fax: +123 (0)123 4445557

Abstract

In addition to vaccines, antiviral drugs are essential for suppressing COVID-19. Although several inhibitor candidates were reported for SARS-CoV-2 main protease, most are highly polar peptidomimetics with poor oral bioavailability and cell membrane permeability. Here, we conducted structure-based virtual screening and *in vitro* assays to obtain hit compounds belonging to a new chemical space excluding secondary amides. In total, 180 compounds were subjected to the primary assay at 20 μM , and nine compounds with inhibition rates higher than 5% were obtained. The IC_{50} of six compounds was determined in dose-response experiments, with the values on the order of 10^{-4} μM . Although nitro groups were enriched in the substructure of the hit compounds, they did not significantly contribute to the binding interaction in

the predicted docking poses. Physicochemical properties prediction showed good oral absorption. These new scaffolds are promising candidates for future optimization.

Introduction

Caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Coronavirus disease 2019 (COVID-19) became a pandemic in 2020 and is still highly prevalent.¹ Although some effective vaccines have been developed²⁻⁴ and are being widely administered,^{5,6} the disease is far from being completely eradicated because of poor compliance by the public with containment protocols, vaccine breakthrough infections, and the emergence of mutant strains.^{7,8} In addition, although antiviral drugs such as remdesivir have shown some efficacy in drug repositioning studies,⁹ no effective and specific SARS-CoV-2 antiviral drugs are available. The most widely used attempts to identify new anti-coronaviral drugs involve targeting RNA-dependent RNA polymerase and main protease (3-chymotrypsin like protease).^{10,11} Main protease is an enzyme that cleaves the viral polyprotein and is essential for viral replication.¹² It shows glutamine-specific cleavage activity which has not been observed in human proteases,^{13,14} and is highly conserved among coronaviruses such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome,¹⁵ making it a suitable target for drug discovery.¹⁶ The structural formulas of representative inhibitors of main protease, including the earliest and most recent ones, are shown in Fig. 1a. Many of these inhibitors are peptidomimetics. N3 is a substrate-mimicking covalent inhibitor identified in a study of SARS-coronavirus (CoV).¹⁷ This inhibitor covalently binds to a cysteine residue in the active site as a Michael acceptor; however, because of its high polarity, it exhibits low membrane permeability and is not effective *in vivo*.^{18,19} GC376 is a dipeptide-based inhibitor of main protease that was originally developed for treating feline infectious peritonitis and is a broad-spectrum anti-coronaviral drug.²⁰⁻²² Additionally, Pfizer has started clinical trials of oral (PF-07321332) and intravenous (PF-07304814) candidate inhibitors.²³ PF-07321332

was optimized from boceprevir,¹⁹ and a nitrile group was introduced as a covalent warhead to react with the active-site cysteine residue (Fig. 1b). GRL-2420 is a tripeptide-based inhibitor²⁴ that was originally found in a study of SARS-CoV. CVD-0013943 is an inhibitor discovered as part of the COVID Moonshot project,²⁵ an open science challenge to fight the global pandemic.²⁶⁻²⁸ CVD-0013943 is smaller in size than other peptidic inhibitors and was shown to have low toxicity but also low metabolic stability.^{29,30} Various other inhibitors are currently being developed by pharmaceutical companies;³¹ however, some of the structural formulas of these inhibitors have not been disclosed. Protease inhibitors tend to have an amide structure,^{32,33} and many active compounds identified to date are highly polar compounds containing amides, as shown in Fig. 1a. The main protease is an enzyme that functions inside of the virus-infected cell;¹⁶ thus, compounds must penetrate the cell membrane to inhibit the protease. Amide compounds sometimes show low membrane permeability because of their polarity or are degraded by proteases,^{34,35} and structural conversion is necessary in some cases, particularly to ensure oral bioavailability. Therefore, non-amide active compounds should be identified to expand the chemical space of hit compounds and increase the success rate of novel drug discovery. In this study, we created a subset of a screening compound library that excludes the secondary amide structure and conducted a hit search for non-amide compounds using structure-based virtual screening (SBVS), a rational *in silico* physicochemical simulation method. Candidate compounds for the assay were selected by SBVS, which is more useful than ligand-based virtual screening (LBVS) for identifying novel scaffolds.³⁶⁻³⁹ The 180 compounds extracted by SBVS were subjected to enzyme inhibition assays to confirm their activity, and six compounds showing activity were obtained.

Results

in silico screening

The Enamine library (3,341,762 compounds) was filtered into 99,765 compounds to avoid secondary amides. These compounds were ranked by conventional rigid docking simulation⁴⁰ using three Protein Data Bank (PDB) structures as targets (6M0K.PDB,^{41,42} 7JKV.PDB^{24,43} and Mpro-x12073.PDB⁴⁴). According to the docking scores and visual inspection, 180 compounds were selected for *in vitro* assays. The list of 180 compounds is shown in Supplementary Table 1. The PDB ID of the protein structure of the docking target, from which the selection of each compound was derived, is also shown in this table. A comparison of the properties of the 180 compounds evaluated in this study and two sets of known hit compounds is shown in Fig. 2. The two sets consist of ChEMBL registered compounds and COVID Moonshot compounds with sub-micromolar activities. The 180 compounds assayed in this study do not contain secondary amides; however, compounds containing lactams or tertiary amides were not excluded (Fig. 2a). Principal component analysis (PCA) plots of each group of hit compounds based on Morgan fingerprints are shown in Fig. 2b and 2c. Although there was some overlap between the sets in the chemical space, the 180 compounds assayed in this study were generally located in a new space.

Primary assay

The 180 compounds selected by *in silico* screening were examined by an *in vitro* fluorescence assay. The mechanism of the assay system is shown in Fig. 3a. The assay system was validated using GC376²⁰ as a positive control (Fig. 3b). All test compounds were assayed at 20 μ M. Among the 180 compounds, nine compounds showed inhibition rates greater than 5% (Supplementary Table 1).

Dose response experiment

Dose-response experiments were conducted for compounds whose inhibition rates in the primary assay were greater than 5%. The dose-response curves are shown in Figure 3c. The concentrations of the compounds that reduced enzyme activity by 50% (IC_{50}) were determined for six compounds: Z391132396, Z166626994, Z819866548, Z2094146478, Z1159100304, and Z324552662. In the counter assay, no signal interference was detected for the six compounds. The structural formulas of these six hit compounds are shown in Fig. 4a. Fig. 4b shows the positions of these hit compounds in the PCA plot in Fig. 2c.

Redocking of the six hit compounds

As five of the six hit compounds were candidates obtained from docking against Mpro-x12073.PDB, each hit compound was redocked against Mpro-x12073.PDB. A superimposed image of the docking poses of the six compounds (white) is shown in Fig. 4c. Ligand x12073 is colored as yellow. All ligands except for Z324552662 were in positions that roughly overlapped with x12073 and occupied the P1-P2 pocket. Z324552662 protruded into the P1' pocket. The nitro groups faced the outside of the cavity. The 2D diagrams of the docking poses are shown in Supplementary Figure 1.

ADME prediction of compounds

Absorption, distribution, metabolism, and excretion (ADME) predictions for the compounds assayed in this study and sets of known inhibitors were performed using SwissADME⁴⁵ (Supplementary Table 2). The six compounds identified in this study generally satisfied Lipinski's rule of five and were predicted to be orally absorbable; structural alerts include nitro groups, whereas no structures corresponded to pan-assay interference compounds (PAINS; compounds prone to false positives).⁴⁶ IC_{50} values, structural alerts, and cLogP values for each compound are shown in Table 1.

Discussion

Of the 180 compounds assayed, we obtained six compounds with main protease inhibitory activity at high concentrations and for which IC_{50} values could be determined. We identified a tertiary amide compound, a sulfonamide compound, a compound containing a lactam structure, and three compounds without an amide bond. Some hit compounds were weakly reactive, such as a weak electrophile containing a nitro group. As we did not perform a counter assay using other enzymes, the target specificity of these compounds remains unknown. Twelve of the 180 compounds assayed contained nitro groups, and four of the six hit compounds contained nitro groups. Because the percentage of compounds containing nitro groups was enriched, the electrophilic nature of the nitro groups may have conferred reactivity to the cysteine protease. Formation of a thiohydroximate adduct in the reaction of a nitro group with a cysteine residue in the active site has been reported.⁴⁷ In fragment screening conducted prior to the COVID Moonshot project,⁴⁸ an electrophile library^{49,50} oriented toward covalent bonding was used. As these electrophile libraries did not cover nitro groups, they evaluated a different chemical space from that examined in the present study. Based on the predicted docking poses (Supplementary Figure 1), covalent interactions were not suggested by our results because the nitro groups of each hit compound were not located close to the cysteine residues in the active site. In addition, because these nitro groups are not responsible for strong interactions in the docking poses, they are considered bioisosterically substitutable⁵¹ or removable if necessary. In the docking poses, the residue interactions were generally consistent with important hot spots reported previously.⁵² All hit compounds showed IC_{50} values on the order of 10^{-4} M, which is weaker than those of currently known sets of amide compounds with sub-micromolar activity. On a case-by-case basis, hit-to-lead optimization can increase activity by hundreds or thousands of fold.⁵³ For example, the activity of an amide bond can be improved by restricting the dihedral degree of freedom of the amide bond to the active conformation by cyclization in some cases.⁵⁴ However, because the structure of the active site of the main protease widely fluctuates,⁵⁵

it may be desirable for the compound to have some degree of freedom in its conformers to accommodate fluctuations in the active site. In contrast, compounds without a secondary amide structure are thought to be more stable against cleavage by proteases, and structural optimization may yield a more stable and active inhibitor *in vivo*.⁵⁶ As our hit compounds did not contain a secondary amide at the P1-P2 position (Fig. 4c), the hit compound structures may be useful as reference scaffolds if amide substitution is necessary to optimize other peptidomimetic inhibitors. In addition, most assay data sets reported to date for main proteases, both positive and negative, consist of compounds containing secondary amide structures; thus, our dataset is valuable because it expands the compound chemical space to be assayed. Some inhibitors are currently being evaluated in clinical trials²³ but their clinical usefulness remains unclear. Expanding the number of hit compounds with higher structural diversity is beneficial for drug discovery. Owing to their predictedly good physicochemical properties as oral drugs, our new scaffolds identified in this study will contribute to the advancement of anti-coronaviral drug research.

Experimental

Filtering of screening compound library

The September 2020 version of the Enamine Collection,⁵⁷ consisting of 3,341,762 compounds, was used as the compound library. To obtain a set of compounds without secondary amide bonds, the library was filtered using the following criteria.

- $20 \leq \text{heavy atom count} \leq 30$
- $1 \leq \text{hydrogen bond donor (HBD) count}$
- $2 \leq \text{hydrogen bond acceptor (HBA) count} \leq 7$
- $0.1 \leq \text{fraction of sp}^3 \text{ carbons (Fsp}^3\text{)} \leq 0.45$

- $2 \leq$ rotatable bond (RB) count ≤ 6
- $2 \leq$ aromatic ring count ≤ 3
- $1 \leq$ aromatic heterocycle count
- $1 \leq$ aliphatic ring count ≤ 2
- (aromatic ring count + aliphatic ring count) ≤ 4
- $300 <$ molecular weight < 450
- contains no amide bond except tertiary amide

After filtering, 99,765 compounds remained.

Structure-based virtual screening

Compound conformer generation

After filtering as described above, conformer generation was performed for the 99,765 remaining compounds using GYPSUM-DL software (version 1.1.7).⁵⁸ The following execution options were used: `--max_variants_per_compound 1 --use_durrant_lab_filters`

The resulting structures were saved as SDF files.

Protein model preparation

The protein models for docking simulation were prepared using Molegro Virtual Docker (version 7.0.0).⁴⁰ The source PDB structures were 6M0K.PDB,⁴² 7JKV.PDB⁴³ and Mpro-x12073.PDB⁴⁴ (COVID Moonshot project²⁸). Models were prepared using the Protein Preparation Wizard in Molegro Virtual Docker (default settings).

Compound selection by docking simulation

Using the compound conformers and protein models, docking simulation was performed using Molegro Virtual Docker.⁴⁰ The search space was set as an 8 angstrom sphere centered at the active site. Docking simulation was performed via the Docking Wizard in Molegro Virtual Docker (Scoring function: PLANTS score, Algorithm: MolDock SE, After docking: Energy Minimization enabled, H-bonds optimization enabled. The compounds were ranked by the Rerank score (linear combination of steric, Van der Waals, hydrogen bonding and electrostatic interactions) and LE3 score (Rerank score divided by heavy atom count). A 2D scatter plot was drawn using the Rerank score and LE3, and we manually chose compounds that were outliers with better (lower) values in the distribution. High-scoring compounds were further assessed by visual inspection to check for key hydrogen bonds and shape fitting. Briefly, docking poses fixed by hydrogen bonds at both ends or three or more points of the compound in the cavity of the active site were selected. Docking poses with geometric centers too close to the walls of the cavity and showing a low filling rate of the cavity were avoided. We also avoided compounds with unfavorable torsions in the conformer of the docked pose. Finally, 180 compounds were chosen for *in vitro* analysis (Supplementary Table 1).

3CL protease in vitro fluorescence assay

Briefly, SARS-CoV-2 3CL protease (0.6 ng/mL) was incubated with fluorogenic 3CL substrate (40 μ M) and test compound (20 μ M) in 25 μ L of assay buffer for 4 h at 25 °C. This experiment was performed by Bienta (Biology Service Division of Enamine Ltd., Kyiv, Ukraine).

Materials

SARS-CoV-2 3CL protease (untagged) (Cat #100823), internally quenched fluorogenic (FRET) 3CL protease substrate (Cat #79952), 3CL protease assay buffer (Cat #79956), and reference protease inhibitor GC376 (Cat #78013) were purchased from BPS Bioscience (San Diego,

CA, USA). In addition, 384-well low-volume black polystyrene microplates with nonbinding surfaces were obtained from Corning (Item #4514; Corning, NY, USA).

Primary assay protocol

In one 384-well plate, 150 nL of compounds was placed in columns 3-22 and dimethyl sulfoxide was added to columns 1-2, except for wells A1-2 and B1-2, which were filled with the reference protease inhibitor GC-376 at the IC_{50} . Columns 23 and 24 were filled with 150 nL of the reference compound GC-376 at a saturating concentration of 3.7 μ M. Next, 7.5 μ L of 3CL protease 2x solution (4.5 μ g/mL in 1x assay buffer with 1 mM DDT) was added to all wells using a Thermo Fisher Scientific Multidrop Combi Reagent Dispenser (Waltham, MA, USA). The enzyme was preincubated with the compounds for 30 min at room temperature (25 °C) with slow shaking. Each well of the plate was filled with 7.5 μ L of 3CL substrate 2x solution (30 μ M in assay buffer with 1 mM DDT) using a Thermo Fisher Scientific Multidrop Combi Dispenser. The final concentration of test compounds was 20 μ M. The plate was incubated for 20 min at room temperature. The fluorescence (excitation 360 nm, emission 460 nm) was read on a Paradigm reader (Molecular Devices, Sunnyvale, CA, USA).

Data analysis of primary assay

Each high-throughput screening plate contained a single test compound in columns 3-22, controls (enzyme, no compound) in columns 1 and 2, and blanks (saturating concentration of the reference compound GC-376) in columns 23 and 24. The high-throughput screening percent inhibition was calculated for each compound from the signal in fluorescence units, mean of the plate controls, and mean of the plate blanks using the following equation: % Inhibition = $100 \times (1 - ((\text{signal-blank mean}) / (\text{control mean-blank mean})))$.

Dose-response analysis

Using the same protocol as used in the primary assay, six compounds were titrated by 3-fold in 8-point curves from 900 μM ($n = 4$). The dose-response curve data were analyzed in GraphPad Prism 8.0.2 software (GraphPad, Inc., La Jolla, CA, USA) and the IC_{50} values were determined by curve fitting.

Author contributions statement

MS conceived the project, KY conducted the *in silico* screening. KY wrote the manuscript. All authors reviewed the manuscript.

Competing interests

There is no conflict of interest to disclose.

Acknowledgement

This work was partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP20am0101112 and the Japanese Society for the Promotion of Science (JSPS) KAKENHI under Grant Number 20H00620.

Supporting Information Available

References

- (1) World Health Organization, WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int/>, Accessed: 2021-5-5.

- (2) Centers for Disease Control and Prevention, Different COVID-19 Vaccines. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html>, 2021; Accessed: 2021-5-5.
- (3) Polack, F. P. et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **2020**, *383*, 2603–2615.
- (4) Baden, L. R. et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **2021**, *384*, 403–416.
- (5) Lab, G. C. D. Coronavirus (COVID-19) Vaccinations. <https://ourworldindata.org/covid-vaccinations>, Accessed: 2021-5-5.
- (6) Dagan, N.; Barda, N.; Kepten, E.; Miron, O.; Perchik, S.; Katz, M. A.; Hernán, M. A.; Lipsitch, M.; Reis, B.; Balicer, R. D. BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. *N. Engl. J. Med.* **2021**,
- (7) Hacısuleyman, E.; Hale, C.; Saito, Y.; Blachere, N. E.; Bergh, M.; Conlon, E. G.; Schaefer-Babajew, D. J.; DaSilva, J.; Muecksch, F.; Gaebler, C.; Lifton, R.; Nussenzweig, M. C.; Hatzioannou, T.; Bieniasz, P. D.; Darnell, R. B. Vaccine Breakthrough Infections with SARS-CoV-2 Variants. *N. Engl. J. Med.* **2021**,
- (8) Kustin, T. et al. *bioRxiv*.
- (9) Beigel, J. H. et al. Remdesivir for the Treatment of Covid-19 - Final Report. *N. Engl. J. Med.* **2020**, *383*, 1813–1826.
- (10) Zumla, A.; Chan, J. F. W.; Azhar, E. I.; Hui, D. S. C.; Yuen, K.-Y. Coronaviruses - drug discovery and therapeutic options. *Nat. Rev. Drug Discov.* **2016**, *15*, 327–347.
- (11) Gil, C.; Ginex, T.; Maestro, I.; Nozal, V.; Barrado-Gil, L.; Cuesta-Geijo, M. Á.; Urquiza, J.; Ramírez, D.; Alonso, C.; Campillo, N. E.; Martinez, A. COVID-19: Drug Targets and Potential Treatments. *J. Med. Chem.* **2020**, *63*, 12359–12386.

- (12) Rota, P. A. et al. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* **2003**, *300*, 1394–1399.
- (13) Goetz, D. H.; Choe, Y.; Hansell, E.; Chen, Y. T.; McDowell, M.; Jonsson, C. B.; Roush, W. R.; McKerrow, J.; Craik, C. S. Substrate Specificity Profiling and Identification of a New Class of Inhibitor for the Major Protease of the SARS Coronavirus. *Biochemistry* **2007**, *46*, 8744–8752.
- (14) Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; de Wilde, A.; Snijder, E. J.; Liu, H.; Hilgenfeld, R. α -Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **2020**, *63*, 4562–4578.
- (15) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595–6628.
- (16) Ullrich, S.; Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127377.
- (17) Yang, H. et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **2005**, *3*, e324.
- (18) Jin, Z. et al. Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293.
- (19) Westberg, M.; Su, Y.; Zou, X.; Ning, L.; Hurst, B.; Tarbet, B.; Lin, M. Z. *bioRxiv*.
- (20) Sharun, K.; Tiwari, R.; Dhama, K. Protease inhibitor GC376 for COVID-19: Lessons learned from feline infectious peritonitis. *Ann Med Surg (Lond)* **2021**, *61*, 122–125.

- (21) Vuong, W.; Khan, M. B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H. A.; McKay, R. T.; van Belkum, M. J.; Joyce, M. A.; Young, H. S.; Lorne Tyrrell, D.; Vederas, J. C.; Joanne Lemieux, M. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* **2020**, *11*, 1–8.
- (22) Fu, L. et al. Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* **2020**, *11*, 1–8.
- (23) Pfizer unveils its oral SARS-CoV-2 inhibitor. <https://cen.acs.org/acs-news/acs-meeting-news/Pfizer-unveils-oral-SARS-CoV/99/i13>, Accessed: 2021-5-5.
- (24) Hattori, S.-I. et al. A small molecule compound with an indole moiety inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* **2021**, *12*, 668.
- (25) Chodera, J. D. THE COVID MOONSHOT -Closing in on an orally-bioavailable small molecule inhibitor of SARS-CoV-2 Mpro through a global open science collaboration. Winter RosettaCon, 2021.
- (26) Kantsadi, A. L.; Cattermole, E.; Matsoukas, M.-T.; Spyroulias, G. A.; Vakonakis, I. *Cold Spring Harbor Laboratory*.
- (27) PostEra, COVID Moonshot: An international effort to DISCOVER A COVID ANTIVIRAL. <https://covid.postera.ai/covid>, Accessed: 2021-3-31.
- (28) The COVID Moonshot Consortium, et al. *bioRxiv*.
- (29) Twitter@covid_moonshot, COVID Moonshot statement. https://twitter.com/covid_moonshot/status/1387372271069581315?s=20, 2021; Accessed: 2021-5-5.
- (30) Chodera, J. D. THE COVID MOONSHOT -An open science collaboration to develop an orally bioavailable inhibitor of SARS-CoV-2 main viral protease. ACP Spring, 2021.

- (31) Heptares, S. SOSEI HEPTARES CORONAVIRUS PROGRAM — NOV 2020. 2020.
- (32) Drag, M.; Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* **2010**, *9*, 690–701.
- (33) Otto, H.-H.; Schirmeister, T. Cysteine Proteases and Their Inhibitors. *Chem. Rev.* **1997**, *97*, 133–172.
- (34) Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem. Biol.* **2014**, *21*, 1115–1142.
- (35) T Borchardt R.; Jeffrey, A.; Siahaan, T. J.; Gangwar, S.; Pauletti, G. M. Improvement of oral peptide bioavailability: Peptidomimetics and prodrug strategies. *Adv. Drug Deliv. Rev.* **1997**, *27*, 235–256.
- (36) Broccatelli, F.; Brown, N. Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. *J. Chem. Inf. Model.* **2014**, *54*, 1634–1641.
- (37) Chiba, S. et al. An iterative compound screening contest method for identifying target protein inhibitors using the tyrosine-protein kinase Yes. *Sci. Rep.* **2017**, *7*, 1–13.
- (38) Chiba, S. et al. Identification of potential inhibitors based on compound proposal contest: Tyrosine-protein kinase Yes as a target. *Sci. Rep.* **2015**, *5*, 1–13.
- (39) Chiba, S. et al. A prospective compound screening contest identified broader inhibitors for Sirtuin 1. *Sci. Rep.* **2019**, *9*, 1–12.
- (40) Thomsen, R.; Christensen, M. H. MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49*, 3315–3321.
- (41) Dai, W. et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368*, 1331–1335.

- (42) Zhang, B., Zhao, Y., Jin, Z., Liu, X., Yang, H., Liu, H., Rao, Z., Jiang, H., 6M0K.PDB: The crystal structure of COVID-19 main protease in complex with an inhibitor 11b. Title of the publication associated with this dataset: Protein Data Bank.
- (43) Bulut, H., Hattori, S.I., Das, D., Murayama, K., Mitsuya, H., 7JKV.PDB: Crystal Structure of SARS-CoV-2 main protease in complex with an inhibitor GRL-2420. Title of the publication associated with this dataset: A small molecule compound with an indole moiety inhibits the main protease of SARS-CoV-2 and blocks virus replication.
- (44) Diamond, X. . Mpro-x12073.PDB. <https://fragalysis.diamond.ac.uk/viewer/react/preview/direct/target/Mpro/mols/x12073/L/P/C>, Accessed: 2021-5-5.
- (45) Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717.
- (46) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.
- (47) Ray, S.; Kreitler, D. F.; Gulick, A. M.; Murkin, A. S. The Nitro Group as a Masked Electrophile in Covalent Enzyme Inhibition. *ACS Chem. Biol.* **2018**, *13*, 1470–1473.
- (48) Douangamath, A. et al. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **2020**, *11*, 1–11.
- (49) Keeley, A.; Ábrányi-Balogh, P.; Keserű, G. M. Design and characterization of a heterocyclic electrophilic fragment library for the discovery of cysteine-targeted covalent inhibitors. *Med. Chem. Commun.* **2019**, *10*, 263–267.
- (50) Resnick, E. et al. Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *J. Am. Chem. Soc.* **2019**, *141*, 8951–8968.

- (51) Meanwell, N. A. In *Tactics in Contemporary Drug Design*; Meanwell, N. A., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2015; pp 283–381.
- (52) Yoshino, R.; Yasuo, N.; Sekijima, M. Identification of key interactions between SARS-CoV-2 main protease and inhibitor drug candidates. *Sci. Rep.* **2020**, *10*, 1–8.
- (53) Keserü, G. M.; Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discov. Today* **2006**, *11*, 741–748.
- (54) Schärfer, C.; Schulz-Gasch, T.; Ehrlich, H.-C.; Guba, W.; Rarey, M.; Stahl, M. Torsion Angle Preferences in Druglike Chemical Space: A Comprehensive Guide. *J. Med. Chem.* **2013**, *56*, 2016–2028.
- (55) Komatsu, T. S.; Okimoto, N.; Koyama, Y. M.; Hirano, Y.; Morimoto, G.; Ohno, Y.; Taiji, M. Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation. *Sci. Rep.* **2020**, *10*, 1–11.
- (56) Lenci, E.; Trabocchi, A. Peptidomimetic toolbox for drug discovery. *Chem. Soc. Rev.* **2020**, *49*, 3262–3277.
- (57) Screening Collection - Enamine. <https://enamine.net/hit-finding/compound-collections/screening-collection>, Accessed: 2020-9-30.
- (58) Ropp, P. J.; Spiegel, J. O.; Walker, J. L.; Green, H.; Morales, G. A.; Milliken, K. A.; Ringe, J. J.; Durrant, J. D. Gypsum-DL: an open-source program for preparing small-molecule libraries for structure-based virtual screening. *J. Cheminform.* **2019**, *11*, 34.

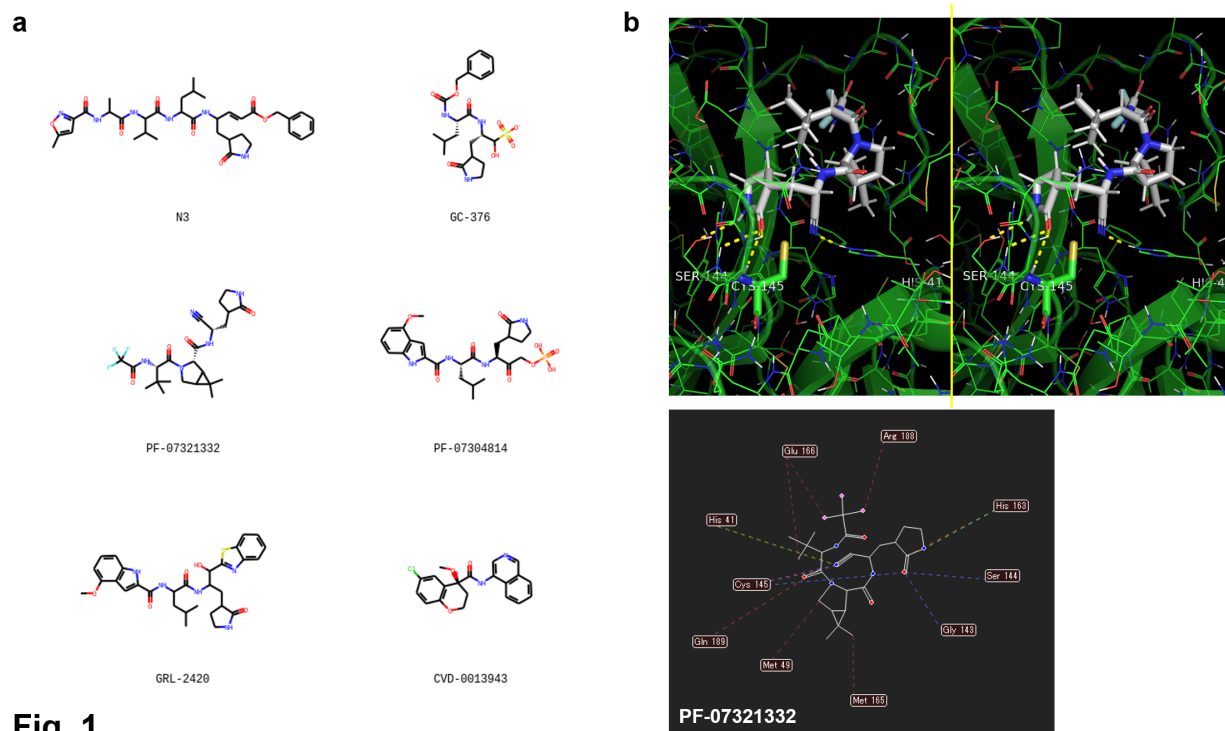


Fig. 1

Figure 1: Examples of known main protease inhibitors.

(a) Structural formulas of inhibitors. (b) Predicted docking pose of PF-07321332 and main protease. The protein structure is 7NBR.PDB. 3D representation is a wall-eye stereogram powered by PyMOL. 2D interaction diagram was output by Molegro Virtual Docker.

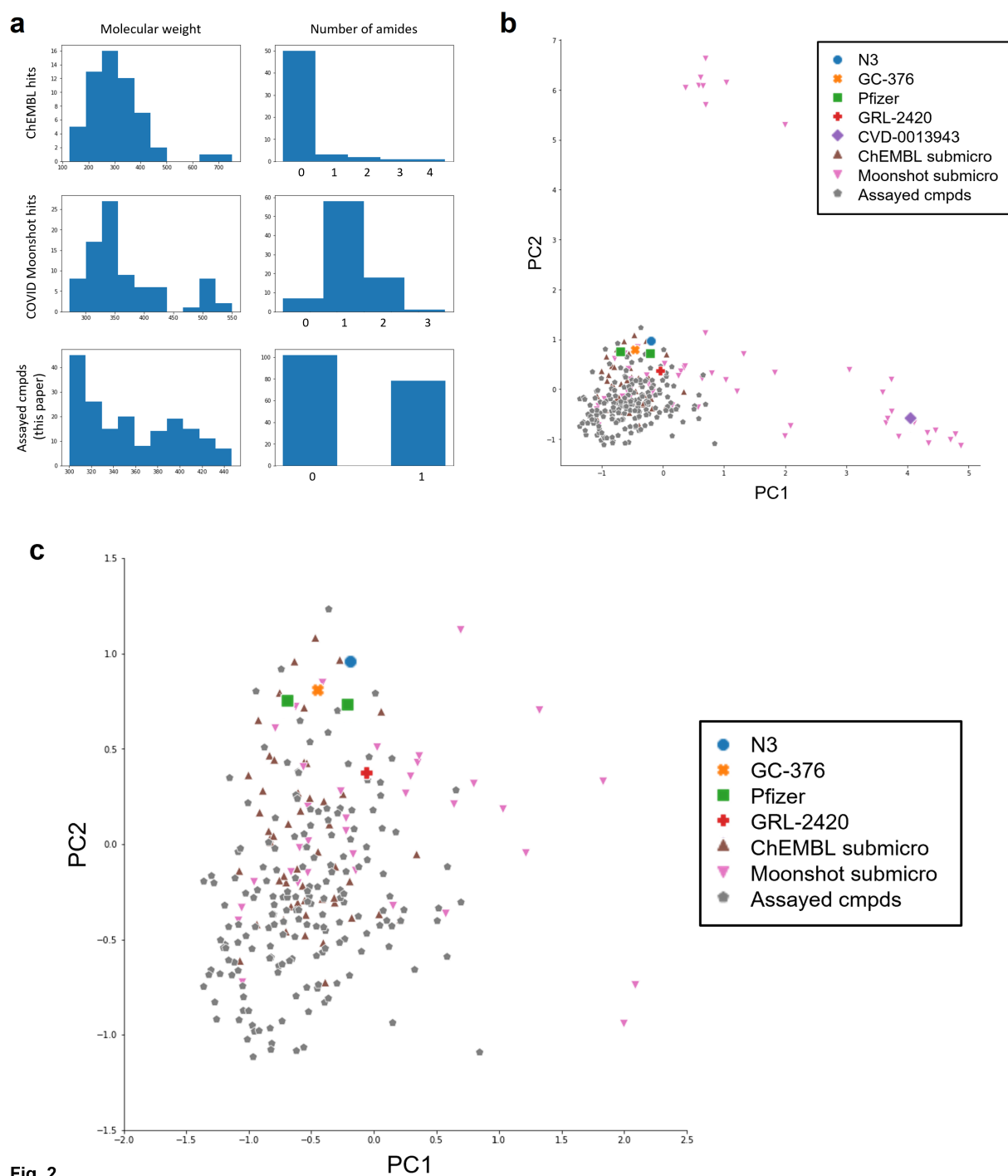


Fig. 2

Figure 2: Comparison of the compound sets.

(a) Distribution of molecular weight and amide number of compounds for each data source. Among the compounds showing inhibitory activity on the main protease, those with $IC_{50} < 1$ μ M were extracted from ChEMBL and COVID Moonshot. (b) PCA plots of chemical space for compounds from each data source and the compounds shown in Fig. 1a. Contribution ratio: PC1 0.0548, PC2 0.0384 (c) Enlargement of the dense part of Fig. 2b.

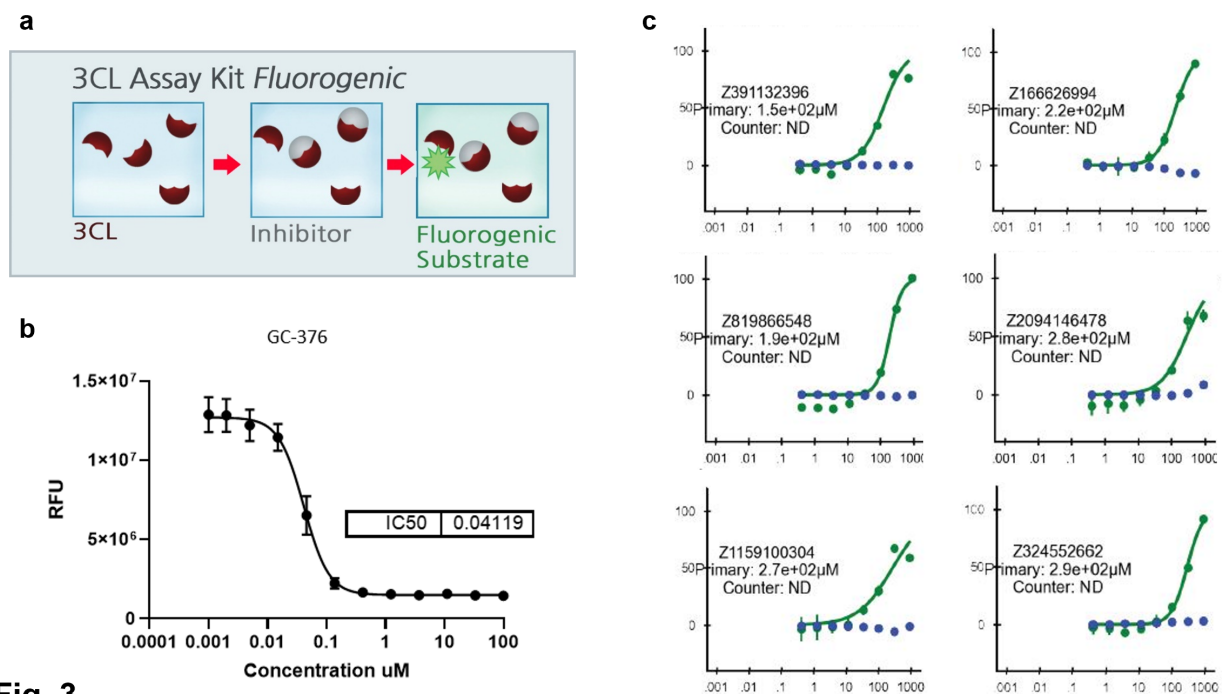


Fig. 3

Figure 3: In vitro assay setup and results.

(a) Schematic of the assay system. The decrease in fluorescence caused by the inhibitor was measured. (b) Validation of the assay system. The dose-response curve of the positive control compound GC-376 was measured. (c) Dose-response curves of six hit compounds whose IC₅₀ were determined by dose-response experiments. The IC₅₀ values are shown along with the graph.

Table 1: Properties of the six hit compounds. IC₅₀ values, structural alerts, and cLogP values for each compound are shown.

ID	IC ₅₀ [μM]	Structural alerts	cLogP (ZINC db)
Z391132396	154	Nitro-Sp2	1.95
Z166626994	222	Nitro-Sp2	2.529
Z819866548	189	Nitro-Sp2	1.864
Z2094146478	281	-	2.877
Z1159100304	273	Nitro-Sp2, Sulfone-Cyclic	1.804
Z324552662	291	Propenals, Alkene-Internal-Sp2	4.066

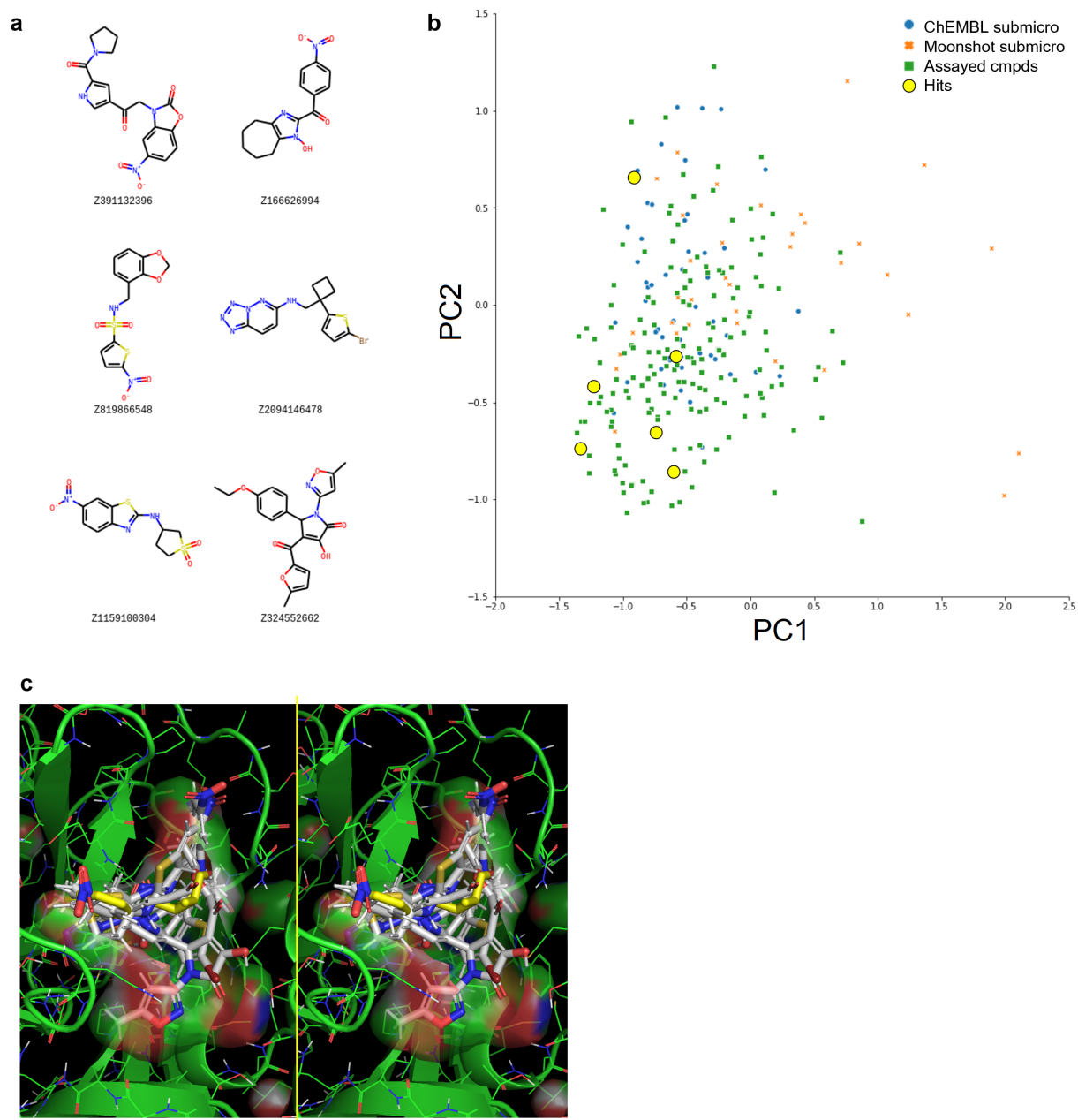


Fig. 4

Figure 4: Hit compounds and their positions. (a) Structural formulas of the six hit compounds. (b) Positions of the six hits in the PCA plot of Fig. 2c. (c) Docking poses of the six hit compounds (white) and ligand x12073 (yellow) superimposed at the active site.